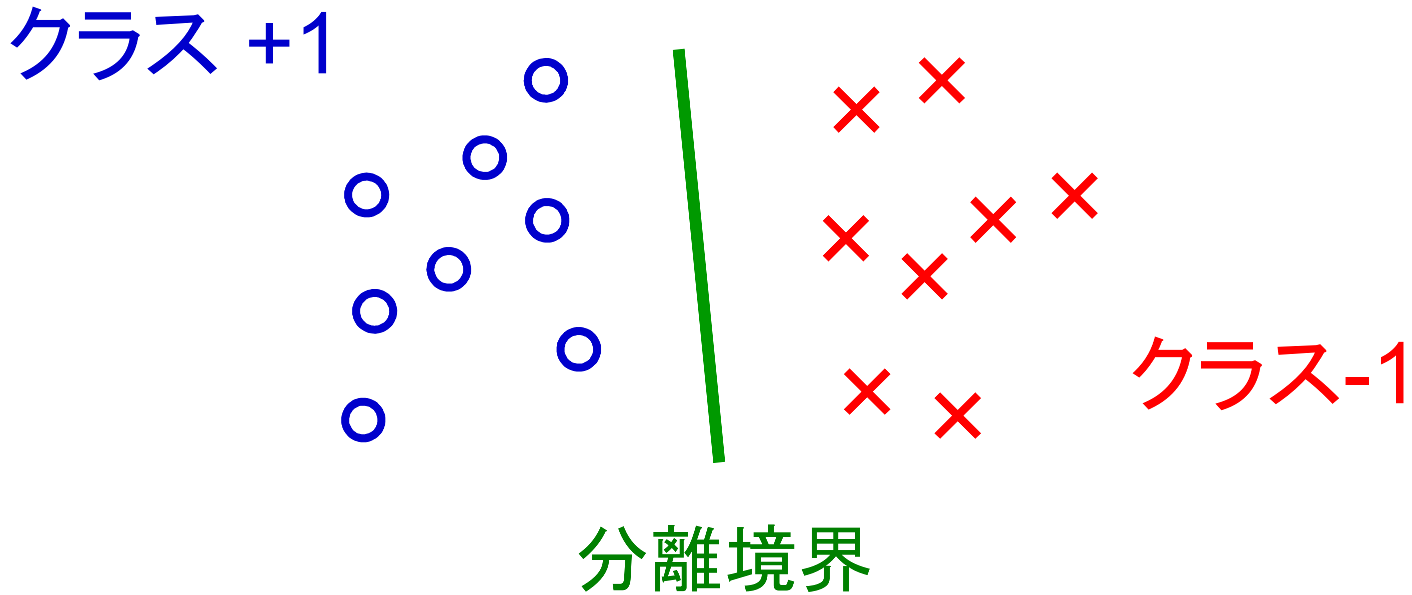
確率的分類（10章）と系列データの分類（11章）

杉山将・本多淳也
sugi@k.u-tokyo.ac.jp, jhonda@k.u-tokyo.ac.jp
http://www.ms.k.u-tokyo.ac.jp

2クラスの分類問題

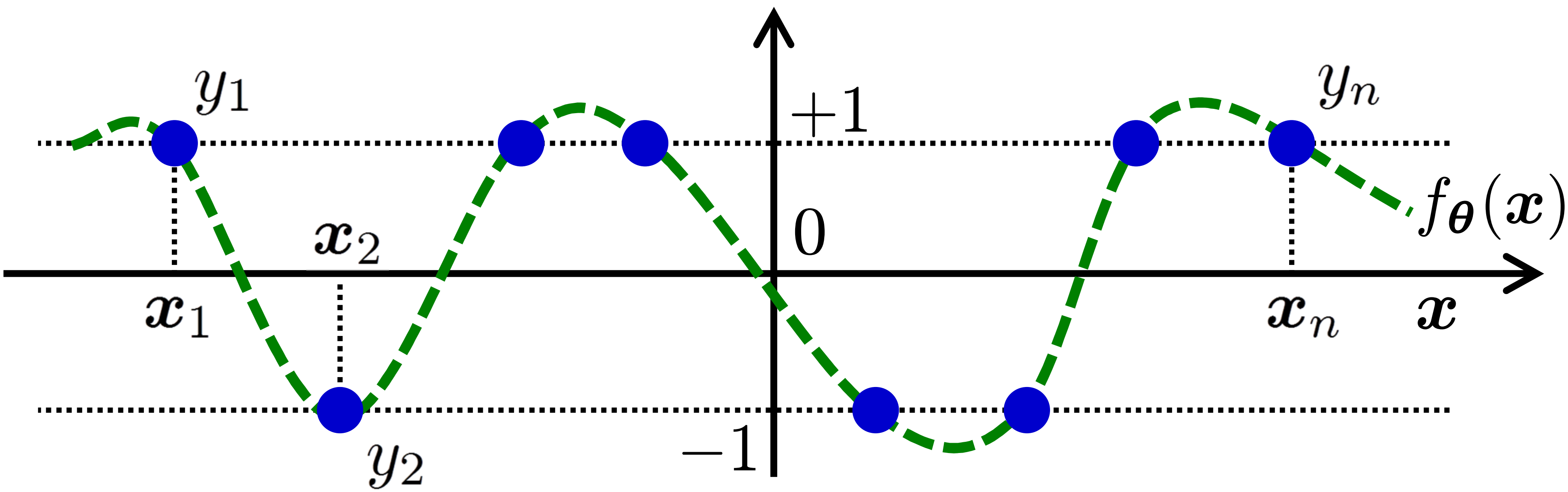
- ラベル付き訓練データ: $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$
  - 入力 $\boldsymbol{x}$ は d 次元の実ベクトル $\boldsymbol{x} \in \mathbb{R}^d$
  - 出力 y は2値のクラスラベル $y \in \{+1, -1\}$



- クラス間の分離境界を求めたい

2クラスの分類問題

- 2クラス分類問題は2値関数の近似問題と等価：



- 回帰学習法が分類にも使える！

パラメータに関する線形モデル

■ 線形モデル: $$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{b} \theta_j \phi_j(\boldsymbol{x})$$

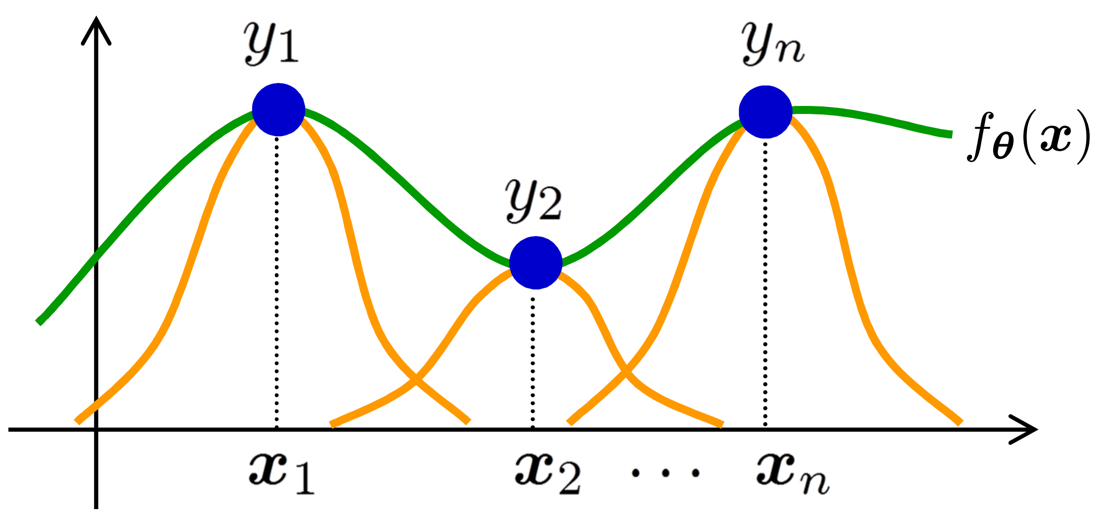
$\{\phi_j(\boldsymbol{x})\}_{j=1}^{b}$:基底関数

■ カーネルモデル:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{n} \theta_j K(\boldsymbol{x}, \boldsymbol{x}_j)$$

ガウスカーネル

$$K(\boldsymbol{x}, \boldsymbol{c}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{c}\|^2}{2h^2}\right)$$



回帰学習による分類

■パラメータを正則化最小二乗回帰で学習

$$\widehat{\boldsymbol{\theta}} = \operatorname*{argmin}_{\boldsymbol{\theta}} \left[\frac{1}{2} \sum_{i=1}^{n} \left(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - y_i \right)^2 + \frac{\lambda}{2} \|\boldsymbol{\theta}\|^2 \right]$$

$\lambda\ (\geq 0)$: 正則化パラメータ

■テストパターンの分類:

$$\widehat{y} = \mathrm{sign}\left(f_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x}) \right) = \begin{cases} +1 & (f_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x}) > 0) \\ 0 & (f_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x}) = 0) \\ -1 & (f_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x}) < 0) \end{cases}$$

0/1-損失関数とマージン

- 分類問題では，学習した関数の符号だけが必要

$$\widehat{y} = \mathrm{sign}\left(f_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x})\right)$$

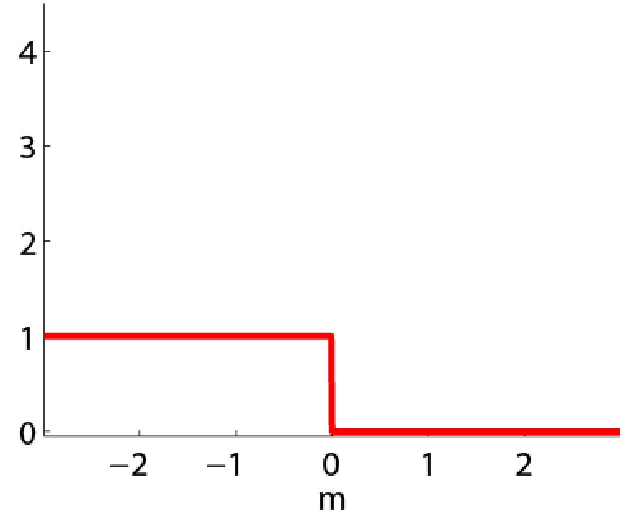
- ℓ_2-損失でなく，0/1-損失の方が自然

$$J_{0/1}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{n}\left(1 - \mathrm{sign}\left(m_i\right)\right)$$

$$m_i = f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)y_i$$

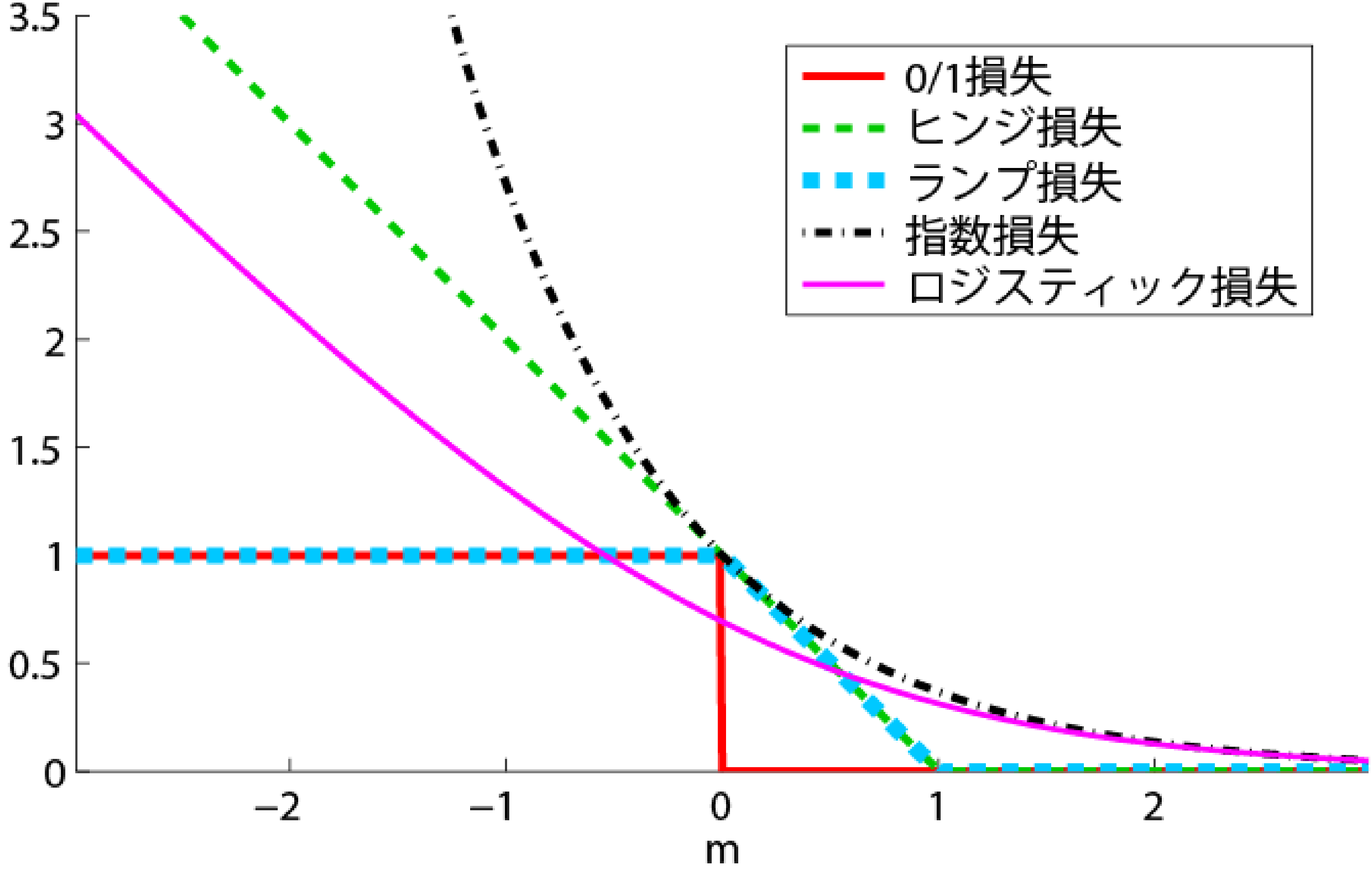
マージン

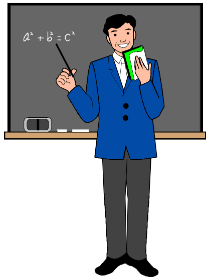
- $J_{0/1}(\boldsymbol{\theta})$ は誤分類標本数に相当
  - 分類の損失としては理想的
  - しかし傾きを持たない離散的な関数
  - 0/1-損失の最小化はNP困難



代理損失

- 分類には様々な代理損失が用いられる





講義の流れ

1. 確率的分類（10章）
    - A) ロジスティック回帰
    - B) 確率的勾配法
    - C) ロジスティック損失最小化
    - D) 二乗誤差最小化
2. 系列データの分類（11章）

確率的分類

- これまで，学習した関数の符号

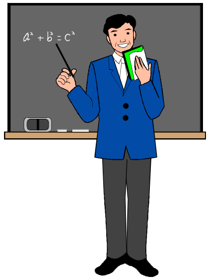
$$\widehat{y} = \mathrm{sign}\left(f_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x})\right)$$

によって分類を行う手法を紹介してきた

- クラス事後確率 $p(y = \widehat{y}|\boldsymbol{x})$ の学習に基づく分類:

$$\widehat{y} = \underset{y=1,\ldots,c}{\mathrm{argmax}}\, p(y|\boldsymbol{x})$$

  - クラス予測に対する信頼度が得られる（信頼度が低い場合は自動分類をあきらめる等）
  - 多クラス分類が自然に行える



講義の流れ

1. 確率的分類（10章）
   - A) ロジスティック回帰
   - B) 確率的勾配法
   - C) ロジスティック損失最小化
   - D) 二乗誤差最小化
2. 系列データの分類（11章）

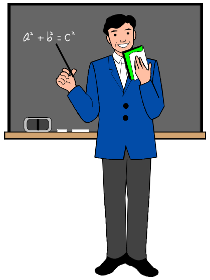
ロジスティック回帰

- ロジスティックモデル:
  - 一般化線形モデルの一種

$$q(y|\boldsymbol{x};\boldsymbol{\theta}) = \frac{\exp\left(\sum_{j=1}^{b}\theta_j^{(y)}\phi_j(\boldsymbol{x})\right)}{\sum_{y'=1}^{c}\exp\left(\sum_{j=1}^{b}\theta_j^{(y')}\phi_j(\boldsymbol{x})\right)}$$

$$\boldsymbol{\theta} = (\underbrace{\theta_1^{(1)},\ldots,\theta_b^{(1)}}_{\text{class } 1},\ldots,\underbrace{\theta_1^{(c)},\ldots,\theta_b^{(c)}}_{\text{class } c})^\top$$

- $(\boldsymbol{\theta}^{(y)})^\top\phi(\boldsymbol{x})$ が大きい y ほど現れやすい



講義の流れ

1. 確率的分類(10章)
   - A) ロジスティック回帰
   - B) 確率的勾配法
   - C) ロジスティック損失最小化
   - D) 二乗誤差最小化
2. 系列データの分類(11章)

数学演習

- ロジスティックモデルに対する対数尤度

$$J_i(\boldsymbol{\theta}) = \log q(y_i|\boldsymbol{x}_i;\boldsymbol{\theta})$$

$$q(y|\boldsymbol{x};\boldsymbol{\theta}) = \frac{\exp\left(\sum_{j=1}^{b}\theta_j^{(y)}\phi_j(\boldsymbol{x})\right)}{\sum_{y'=1}^{c}\exp\left(\sum_{j=1}^{b}\theta_j^{(y')}\phi_j(\boldsymbol{x})\right)}$$

の $(\theta_1^{(y)},\ldots,\theta_b^{(y)})^\top$ に関する偏微分は次式で与えられることを示せ

$$\nabla_y J_i(\boldsymbol{\theta}) = -\frac{\exp\left(\sum_{j=1}^{b}\theta_j^{(y)}\phi_j(\boldsymbol{x}_i)\right)\boldsymbol{\phi}(\boldsymbol{x}_i)}{\sum_{y'=1}^{c}\exp\left(\sum_{j=1}^{b}\theta_j^{(y')}\phi_j(\boldsymbol{x}_i)\right)} + \begin{cases}\boldsymbol{\phi}(\boldsymbol{x}_i) & (y = y_i)\\ \boldsymbol{0} & (y \neq y_i)\end{cases}$$

確率的勾配法

1. パラメータ $\boldsymbol{\theta}$ を適当に初期化
2. 標本$(\boldsymbol{x}_i, y_i)$をランダムに選び，勾配を少し上昇

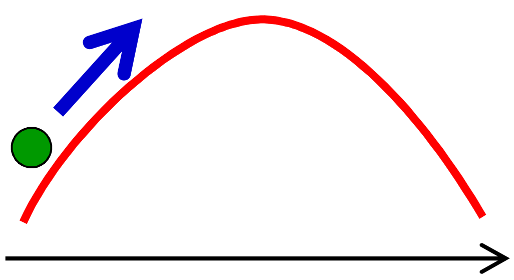
$$\boldsymbol{\theta}^{(y)} \longleftarrow \boldsymbol{\theta}^{(y)} + \varepsilon \nabla_y J_i(\boldsymbol{\theta}) \quad \text{for } y = 1, \ldots, c$$

学習係数

$\varepsilon > 0$

$$\nabla_y J_i(\boldsymbol{\theta}) = -\frac{\exp\left(\boldsymbol{\theta}^{(y)\top} \boldsymbol{\phi}(\boldsymbol{x}_i)\right) \boldsymbol{\phi}(\boldsymbol{x}_i)}{\sum_{y'=1}^{c} \exp\left(\boldsymbol{\theta}^{(y')\top} \boldsymbol{\phi}(\boldsymbol{x}_i)\right)} + \begin{cases} \boldsymbol{\phi}(\boldsymbol{x}_i) & (y = y_i) \\ \mathbf{0} & (y \neq y_i) \end{cases}$$

3. 収束するまで 2. を繰り返す



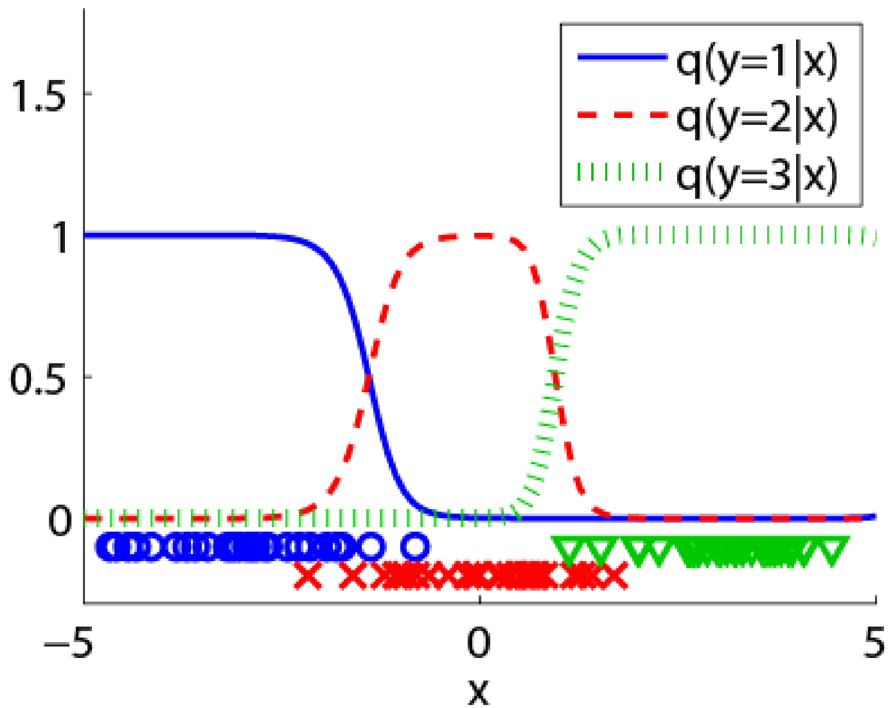
実行例

■ 対数ガウスカーネルモデル

$$q(y|\boldsymbol{x};\boldsymbol{\theta}) \propto \exp\left(\sum_{j=1}^{n}\theta_j K(\boldsymbol{x},\boldsymbol{x}_j)\right)$$

$$K(\boldsymbol{x},\boldsymbol{c}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{c}\|^2}{2h^2}\right)$$

に対するロジスティック回帰の実行例



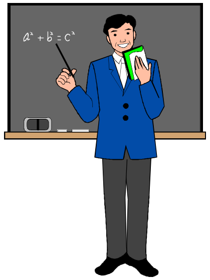
実行例(続き)


```
clear all; rand('state',3); randn('state',3);

n=90; c=3; y=ones(n/c,1)*[1:c]; y=y(:);
x=randn(n/c,c)+repmat(linspace(-3,3,c),n/c,1);
x=x(:);

hh=2*1^2; t0=randn(n,c); e=0.1;
for o=1:n*1000
  i=ceil(rand*n); yi=y(i);
  ki=exp(-(x-x(i)).^2/hh);
  ci=exp(ki'*t0); t=t0-e*(ki*ci)/sum(ci);
  t(:,yi)=t(:,yi)+e*ki;
  if norm(t-t0)<0.000001, break, end
  t0=t;
end
```

次のページに続く

実行例(続き)


```
N=100; X=linspace(-5,5,N)';
K=exp(-(repmat(X.^2,1,n)+repmat(x.^2',N,1)-
2*X*x')/hh);

figure(1); clf; hold on; axis([-5 5 -0.3 1.8]);
C=exp(K*t); C=C./repmat(sum(C,2),1,c);
plot(X,C(:,1),'b-'); plot(X,C(:,2),'r--');
plot(X,C(:,3),'g:');
plot(x(y==1),-0.1*ones(n/c,1),'bo');
plot(x(y==2),-0.2*ones(n/c,1),'rx');
plot(x(y==3),-0.1*ones(n/c,1),'gv');
legend('q(y=1|x)','q(y=2|x)','q(y=3|x)')
```



講義の流れ

1. 確率的分類（１０章）
   - A) ロジスティック回帰
   - B) 確率的勾配法
   - C) ロジスティック損失最小化
   - D) 二乗誤差最小化
2. 系列データの分類（１１章）

数学演習

- 2クラス問題 $y \in \{+1, -1\}$ に対して，ロジスティックモデル

$$q(y|\boldsymbol{x};\boldsymbol{\alpha}) = \frac{\exp\left(\sum_{j=1}^{b} \alpha_j^{(y)} \phi_j(\boldsymbol{x})\right)}{\sum_{y'=\pm 1} \exp\left(\sum_{j=1}^{b} \alpha_j^{(y')} \phi_j(\boldsymbol{x})\right)}$$

は以下と等価であることを示せ

$$q(y|\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{1 + \exp\left(-y f_{\boldsymbol{\theta}}(\boldsymbol{x})\right)}$$

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{b} \theta_j \phi_j(\boldsymbol{x})$$

ロジスティック損失最小化学習

$$q(y|\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{1+\exp\left(-yf_{\boldsymbol{\theta}}(\boldsymbol{x})\right)}$$

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sum_{j=1}^{b} \theta_j \phi_j(\boldsymbol{x})$$

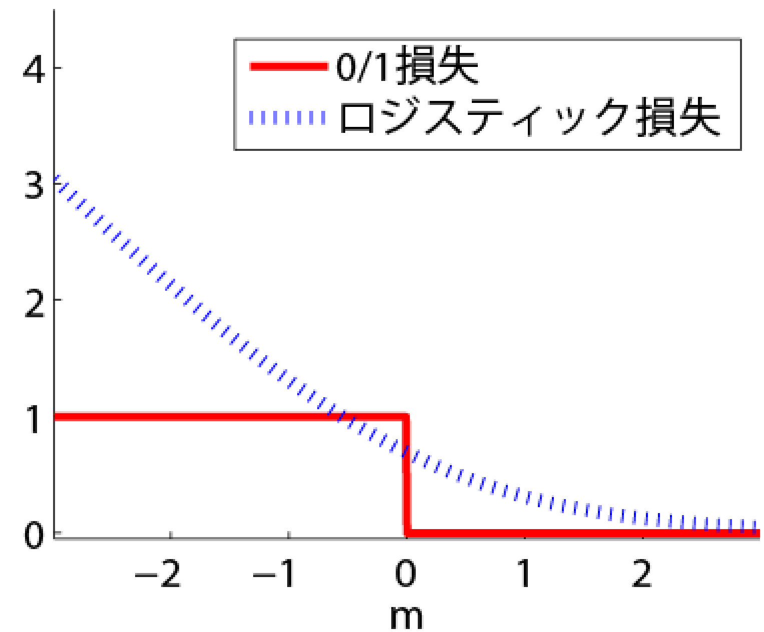
- このロジスティック回帰に対する最尤推定

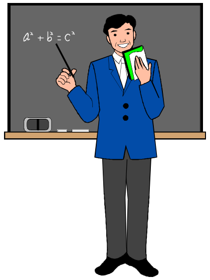
$$\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log q(y_i|\boldsymbol{x}_i;\boldsymbol{\theta})$$

はロジスティック損失
最小化学習と等価

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log\Big(1+\exp(-m_i)\Big)$$

$$m_i = f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) y_i$$





講義の流れ

1. 確率的分類（10章）
   A) ロジスティック回帰
   B) 確率的勾配法
   C) ロジスティック損失最小化
   D) 二乗誤差最小化
2. 系列データの分類（11章）

最小二乗確率的分類

- クラス y の事後確率に対する線形モデル:

$$q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)}) = \sum_{j=1}^{b} \theta_j^{(y)} \phi_j(\boldsymbol{x}) = {\boldsymbol{\theta}^{(y)}}^\top \boldsymbol{\phi}(\boldsymbol{x})$$

  - 必ずしも確率にはならない

- 二乗誤差最小化:

$$J_y(\boldsymbol{\theta}^{(y)}) = \frac{1}{2}\int \Big(q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)}) - p(y|\boldsymbol{x})\Big)^2 p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

  - 尤度最大化の場合は，モデルの正規化が必要
  - 二乗誤差最小化の場合は，正規化しなくてもよい！

二乗誤差の分解

$$J_y(\boldsymbol{\theta}^{(y)}) = \frac{1}{2}\int \Big(q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)}) - p(y|\boldsymbol{x})\Big)^2 p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

$$= \frac{1}{2}\int q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)})^2 p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

$$- \int q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)}) p(y|\boldsymbol{x}) p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

$$+ \frac{1}{2}\int p(y|\boldsymbol{x})^2 p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

定数なので
無視できる

二乗誤差の標本近似

二乗誤差の標本近似:

- $$\int q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)})^2 p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$ ➡ $$\frac{1}{n}\sum_{i=1}^{n} q(y|\boldsymbol{x}_i;\boldsymbol{\theta}^{(y)})^2$$

- $$\int q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)})p(y|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$
$$= \int q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)})p(y)p(\boldsymbol{x}|y)\mathrm{d}\boldsymbol{x}$$
➡ $$\frac{1}{n_y}\sum_{i:y_i=y} q(y|\boldsymbol{x}_i;\boldsymbol{\theta}^{(y)})\frac{n_y}{n}$$

正則化学習規準

■２次正則化項を加えた学習規準:

$$\widehat{J}_y(\boldsymbol{\theta}^{(y)}) = \frac{1}{2n}\sum_{i=1}^{n} q(y|\boldsymbol{x}_i;\boldsymbol{\theta}^{(y)})^2$$

$$-\frac{1}{n}\sum_{i:y_i=y} q(y|\boldsymbol{x}_i;\boldsymbol{\theta}^{(y)}) + \frac{\lambda}{2n}\|\boldsymbol{\theta}^{(y)}\|^2$$

$$= \frac{1}{2n}\boldsymbol{\theta}^{(y)\top}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\boldsymbol{\theta}^{(y)} - \frac{1}{n}\boldsymbol{\theta}^{(y)\top}\boldsymbol{\Phi}^\top\boldsymbol{\pi}^{(y)} + \frac{\lambda}{2n}\|\boldsymbol{\theta}^{(y)}\|^2$$

■最小解は解析的に求まる:

$$\widehat{\boldsymbol{\theta}}^{(y)} = \left(\boldsymbol{\Phi}^\top\boldsymbol{\Phi} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\pi}^{(y)}$$

$$\Phi_{i,j} = \phi_j(\boldsymbol{x}_i)$$

$$\pi_i^{(y)} = \begin{cases} 1 & (y_i = y) \\ 0 & (y_i \neq y) \end{cases}$$

解の補正

$$\widehat{\boldsymbol{\theta}}^{(y)} = \left(\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{\Phi}^\top \boldsymbol{\pi}^{(y)}$$

- 確率は非負で和が1→解を事後補正する

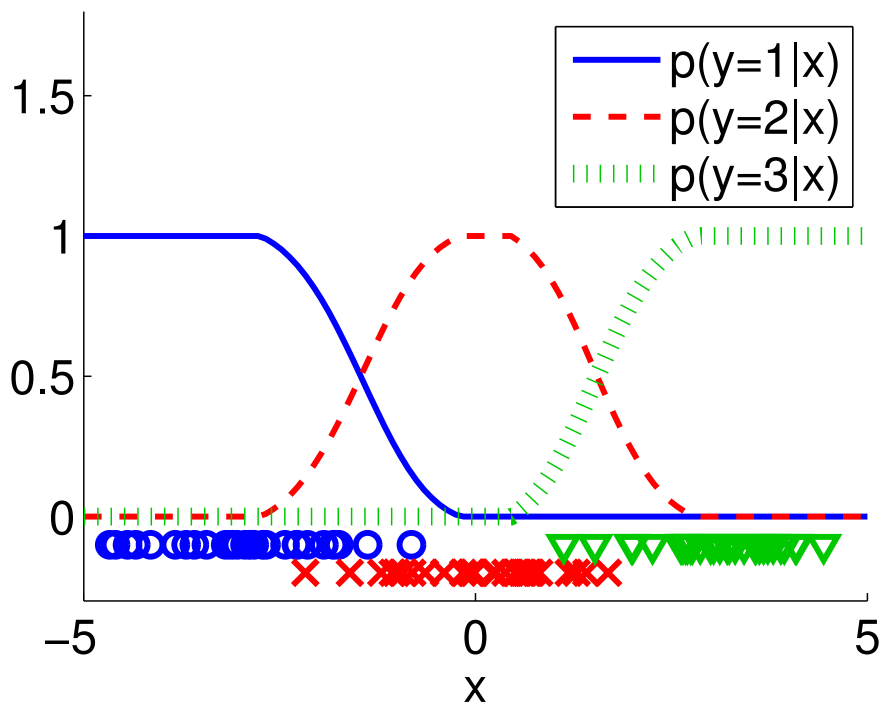
$$\widehat{p}(y|\boldsymbol{x}) = \frac{\max(0, \widehat{\boldsymbol{\theta}}^{(y)\top} \boldsymbol{\phi}(\boldsymbol{x}))}{\sum_{y'=1}^{c} \max(0, \widehat{\boldsymbol{\theta}}^{(y')\top} \boldsymbol{\phi}(\boldsymbol{x}))}$$

実行例

- ガウスカーネルモデル

$$q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)}) = \sum_{j:y_j=y} \theta_j^{(y)} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_j\|^2}{2h^2}\right)$$

に対する実行例

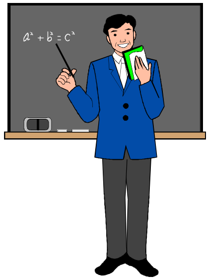


確率的分類：まとめ

- クラス事後確率 $p(y=\widehat{y}|\boldsymbol{x})$ の学習に基づく分類：

$$\widehat{y} = \operatorname*{argmax}_{y=1,\ldots,c} p(y|\boldsymbol{x})$$

  - クラス予測に対する信頼度が得られる（信頼度が低い場合は自動分類をあきらめる等）
  - 多クラス分類が自然に行える
- ロジスティック回帰：
  - 名前は回帰だが，分類学習法
- 最小二乗確率的分類：
  - 計算が簡単



講義の流れ

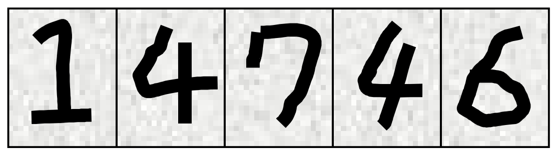
1. 確率的分類（10章）
2. 系列データの分類（11章）
   A) 条件付き確率場モデル
   B) 条件付き確率場モデルの学習
   C) 条件付き確率場モデルによるラベル予測

系列データの認識

$$\overline{\boldsymbol{x}} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)})$$

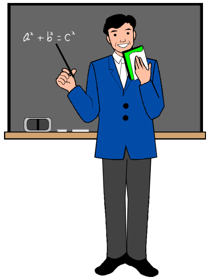
$$\overline{y} = (y^{(1)}, \ldots, y^{(m)})$$

- 例 : 手書き数字列の認識



- 例 : 品詞の分解

こうぎ	が	ねむい



講義の流れ

1. 確率的分類（10章）
2. 系列データの分類（11章）
   - A) 条件付き確率場モデル
   - B) 条件付き確率場モデルの学習
   - C) 条件付き確率場モデルによるラベル予測

系列データの認識

$\overline{\boldsymbol{x}}$	$\boldsymbol{x}^{(1)}$	$\cdots$	$\boldsymbol{x}^{(k-1)}$	$\boldsymbol{x}^{(k)}$	$\cdots$	$\boldsymbol{x}^{(m)}$
$\overline{y}$	$y^{(1)}$	$\cdots$	$y^{(k-1)}$	$y^{(k)}$	$\cdots$	$y^{(m)}$

各データを1つずつ独立に学習・認識:

- ロジスティック回帰をそのまま使えば良い

$$q(y|\boldsymbol{x};\boldsymbol{\theta}) = \frac{\exp\left(\boldsymbol{\theta}_y^\top \boldsymbol{\phi}(\boldsymbol{x})\right)}{\sum_{y'=1}^{c} \exp\left(\boldsymbol{\theta}_{y'}^\top \boldsymbol{\phi}(\boldsymbol{x})\right)}$$

$\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_c^\top)^\top \in \mathbb{R}^{bc}$:パラメータ

- しかし，系列データの特性を生かせない

系列データの認識

$\overline{\boldsymbol{x}}$	$\boldsymbol{x}^{(1)}$	$\cdots$	$\boldsymbol{x}^{(k-1)}$	$\boldsymbol{x}^{(k)}$	$\cdots$	$\boldsymbol{x}^{(m)}$
$\overline{y}$	$y^{(1)}$	$\cdots$	$y^{(k-1)}$	$y^{(k)}$	$\cdots$	$y^{(m)}$

- 全系列データを同時に認識:
  - 系列データの特性を完全に生かせる
  - 学習が困難(クラス数が c^m)

$$q(\overline{y}|\overline{\boldsymbol{x}};\overline{\boldsymbol{\theta}}) = \frac{\exp\left(\overline{\boldsymbol{\theta}}_{\overline{y}}^{\top}\overline{\boldsymbol{\phi}}(\overline{\boldsymbol{x}})\right)}{\sum_{\overline{y}'=1}^{c^m}\exp\left(\overline{\boldsymbol{\theta}}_{\overline{y}'}^{\top}\overline{\boldsymbol{\phi}}(\overline{\boldsymbol{x}})\right)}$$

$\overline{\boldsymbol{\phi}}(\overline{\boldsymbol{x}}) = (\boldsymbol{\phi}(\boldsymbol{x}^{(1)})^{\top}, \ldots, \boldsymbol{\phi}(\boldsymbol{x}^{(m)})^{\top})^{\top} \in \mathbb{R}^{bm}$:基底ベクトル

$\overline{\boldsymbol{\theta}}_{\overline{y}} = (\boldsymbol{\theta}_{\overline{y}}^{(1)\top}, \ldots, \boldsymbol{\theta}_{\overline{y}}^{(m)\top})^{\top} \in \mathbb{R}^{bm}$:パラメータ

条件付き確率場モデル

$\overline{\boldsymbol{x}}$	$\boldsymbol{x}^{(1)}$	$\cdots$	$\boldsymbol{x}^{(k-1)}$	$\boldsymbol{x}^{(k)}$	$\cdots$	$\boldsymbol{x}^{(m)}$
$\overline{y}$	$y^{(1)}$	$\cdots$	$y^{(k-1)}$	$y^{(k)}$	$\cdots$	$y^{(m)}$

- 系列データの特性を生かしつつ，モデルを適度に簡略化する
- 例：隣り合ったデータのみの関連を考慮
  - 連続する2パターンに対応するc^2クラスの認識問題を単に続けて解くのではなく，あくまでもパターン系列全体の認識を同時に行う．

条件付き確率場モデル

$$q(\overline{y}|\overline{\boldsymbol{x}};\boldsymbol{\zeta}) = \frac{\exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y})\right)}{\sum_{\overline{y}'=1}^{\overline{c}} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y}')\right)}$$

- 特徴ベクトルに $\overline{y}$ を含める
- $\boldsymbol{\zeta} = \overline{\boldsymbol{\theta}}$ ， $\boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y}) = \boldsymbol{e}_{\overline{y}}^{(\overline{c})} \otimes \overline{\boldsymbol{\phi}}(\overline{\boldsymbol{x}})$ とおけば，系列全体に対するロジスティックモデルと等価

$$\boldsymbol{e}_{\overline{y}}^{(\overline{c})} = (\underbrace{0,\ldots,0}_{\overline{y}-1},1,\underbrace{0,\ldots,0}_{\overline{c}-\overline{y}})^\top \in \mathbb{R}^{\overline{c}} \qquad \overline{c} = c^m$$

$$(f_1,\ldots,f_u)^\top \otimes \boldsymbol{g} = (f_1\boldsymbol{g}^\top,\ldots,f_u\boldsymbol{g}^\top)^\top$$

- パラメータ数が増え過ぎないよう特徴ベクトルを定める

条件付き確率場モデル

$\overline{\boldsymbol{x}}$	$\boldsymbol{x}^{(1)}$	$\cdots$	$\boldsymbol{x}^{(k-1)}$	$\boldsymbol{x}^{(k)}$	$\cdots$	$\boldsymbol{x}^{(m)}$
$\overline{y}$	$y^{(1)}$	$\cdots$	$y^{(k-1)}$	$y^{(k)}$	$\cdots$	$y^{(m)}$

- 各時刻のラベルの分布がその時刻のパターンとその一つ前のラベルによって定まるモデル:

$$q(\overline{y}|\overline{\boldsymbol{x}};\boldsymbol{\zeta}) = \prod_{k=1}^{m} q(y^{(k)}|\boldsymbol{x}^{(k)}, y^{(k-1)};\boldsymbol{\zeta}) \qquad y^{(0)} = y^{(1)}$$

$$\propto \prod_{k=1}^{m} \exp(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}^{(k)}, y^{(k)}, y^{(k-1)}))$$

$$= \exp\left(\boldsymbol{\zeta}^\top \sum_{k=1}^{m} \boldsymbol{\varphi}(\boldsymbol{x}^{(k)}, y^{(k)}, y^{(k-1)})\right)$$

条件付き確率場モデル

$\overline{\boldsymbol{x}}$	$\boldsymbol{x}^{(1)}$	$\cdots$	$\boldsymbol{x}^{(k-1)}$	$\boldsymbol{x}^{(k)}$	$\cdots$	$\boldsymbol{x}^{(m)}$
$\overline{y}$	$y^{(1)}$	$\cdots$	$y^{(k-1)}$	$y^{(k)}$	$\cdots$	$y^{(m)}$

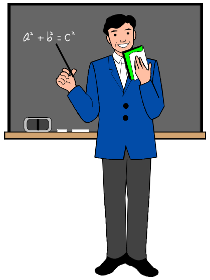
- 各時刻ごとの特徴を足し合わせるモデルに対応：

$$\boldsymbol{\varphi}(\overline{\boldsymbol{x}}, \overline{y}) = \sum_{k=1}^{m} \boldsymbol{\varphi}(\boldsymbol{x}^{(k)}, y^{(k)}, y^{(k-1)})$$

- 例：$(\boldsymbol{\phi}(\boldsymbol{x}^{(k)}), \boldsymbol{e}^{(c)}_{y^{(k-1)}}) \in \mathbb{R}^{b+c}$ を改めて特徴としたロジスティックモデル

$$\boldsymbol{\varphi}(\boldsymbol{x}^{(k)}, y^{(k)}, y^{(k-1)}) = \boldsymbol{e}^{(c)}_{y^{(k)}} \otimes (\boldsymbol{\phi}(\boldsymbol{x}^{(k)}), \boldsymbol{e}^{(c)}_{y^{(k-1)}}) \in \mathbb{R}^{c(b+c)}$$

$$\boldsymbol{e}^{(c)}_{y} = (\underbrace{0, \ldots, 0}_{y-1}, 1, \underbrace{0, \ldots, 0}_{c-y})^{\top} \in \mathbb{R}^{c}$$



講義の流れ

1. 確率的分類（10章）
2. 系列データの分類（11章）
   A) 条件付き確率場モデル
   B) 条件付き確率場モデルの学習
   C) 条件付き確率場モデルによるラベル予測

条件付き確率場モデルの学習

系列データ

$$\left\{(\overline{\boldsymbol{x}}_i, \overline{y}_i) \;\middle|\; \overline{\boldsymbol{x}}_i = (\boldsymbol{x}_i^{(1)}, \ldots, \boldsymbol{x}_i^{(m_i)}),\ \overline{y}_i = \sum_{k=1}^{m_i} c^{k-1}(y_i^{(k)} - 1) + 1\right\}_{i=1}^{n}$$

からパラメータを**最尤推定**により学習：

$$\max_{\boldsymbol{\zeta}} \sum_{i=1}^{n} \log \frac{\exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}_i)\right)}{\sum_{\overline{y}'=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}')\right)}$$

確率的勾配法でパラメータを最適化：

$$\boldsymbol{\zeta} \longleftarrow \boldsymbol{\zeta} + \varepsilon \left(\boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}_i) - \frac{\sum_{\overline{y}=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y})\right) \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y})}{\sum_{\overline{y}'=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}')\right)} \right)$$

条件付き確率場モデルの学習

$$\boldsymbol{\zeta} \longleftarrow \boldsymbol{\zeta} + \varepsilon \left(\boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}_i) - \frac{\sum_{\overline{y}=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}) \right) \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y})}{\sum_{\overline{y}'=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}') \right)} \right)$$

- $\sum_{\overline{y}=1}^{\overline{c}_i}$ の計算に $\mathcal{O}(c^{m_i})$ 時間かかる
  →分母と分子をそれぞれ動的計画法で計算
  - 分母: $\sum_{\overline{y}'=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}') \right)$
  - 分子: $\sum_{\overline{y}=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}) \right) \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y})$

動的計画法による分母の計算

$$\sum_{\overline{y}'=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y}')\right)$$

- $y^{(1)}, \ldots, y^{(m_i-1)}$ と $y^{(m_i)}$ に分解:

$$\sum_{y^{(1)},\ldots,y^{(m_i)}=1}^{c} \exp\left(\sum_{k=1}^{m_i} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)})\right) = \sum_{y^{(m_i)}=1}^{c} A_{m_i}(y^{(m_i)})$$

$$A_\tau(y) = \sum_{y^{(1)},\ldots,y^{(\tau-1)}=1}^{c} \exp\left(\sum_{k=1}^{\tau-1} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)}) + \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(\tau)}, y, y^{(\tau-1)})\right)$$

数学演習

$$A_\tau(y) = \sum_{y^{(1)},\ldots,y^{(\tau-1)}=1}^{c} \exp\left(\sum_{k=1}^{\tau-1} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)}) + \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(\tau)}, y, y^{(\tau-1)})\right)$$

- 以下の再帰表現を求めよ

$$A_\tau(y^{(\tau)}) = \sum_{y^{(\tau-1)}=1}^{c} A_{\tau-1}(y^{(\tau-1)}) \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(\tau)}, y^{(\tau)}, y^{(\tau-1)})\right)$$

$A_\tau(y)$	$\tau = 1$	$\tau = 2$	$\cdots$	$\tau = m$
$y = 1$				
$\vdots$				
$y = c$				

動的計画法による分子の計算

$$\sum_{\overline{y}=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y})\right) \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y})$$

- $y^{(1)}, \ldots, y^{(k'-2)}$ と $y^{(k'-1)}, y^{(k')}$ と $y^{(k'+1)}, \ldots, y^{(m_i)}$ に分解:

$$\sum_{y^{(1)},\ldots,y^{(m_i)}=1}^{c} \exp\left(\sum_{k=1}^{m_i} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)})\right) \left(\sum_{k'=1}^{m_i} \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k')}, y^{(k')}, y^{(k'-1)})\right)$$

$$= \sum_{k'=1}^{m_i} \sum_{y^{(1)},\ldots,y^{(k'-2)}=1}^{c} \sum_{y^{(k'-1)},y^{(k')}=1}^{c} \sum_{y^{(k'+1)},\ldots,y^{(m_i)}=1}^{c}$$

$$\exp\left(\sum_{k=1}^{m_i} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)})\right) \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k')}, y^{(k')}, y^{(k'-1)})$$

- $k < k' - 1$ に関する項と $k > k'$ に関する項はexpの中でのみ現れるのでそれらを $A_{k'-1}$ と $B_{k'}$ で表現

動的計画法による分子の計算

$$\sum_{\overline{y}=1}^{\overline{c}_i} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y})\right) \boldsymbol{\varphi}(\overline{\boldsymbol{x}}_i, \overline{y})$$

- $y^{(1)}, \ldots, y^{(k'-2)}$ と $y^{(k'-1)}, y^{(k')}$ と $y^{(k'+1)}, \ldots, y^{(m_i)}$ に分解:

$$\sum_{y^{(1)},\ldots,y^{(m_i)}=1}^{c} \exp\left(\sum_{k=1}^{m_i} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)})\right) \left(\sum_{k'=1}^{m_i} \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k')}, y^{(k')}, y^{(k'-1)})\right)$$

$$= \sum_{k'=1}^{m_i} \sum_{y^{(1)},\ldots,y^{(k'-2)}=1}^{c} \sum_{y^{(k'-1)},y^{(k')}=1}^{c} \sum_{y^{(k'+1)},\ldots,y^{(m_i)}=1}^{c}$$

$$\exp\left(\sum_{k=1}^{m_i} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)})\right) \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k')}, y^{(k')}, y^{(k'-1)})$$

$$= \sum_{k'=1}^{m_i} \sum_{y^{(k'-1)},y^{(k')}=1}^{c} \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k')}, y^{(k')}, y^{(k'-1)})$$

$$\times \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k')}, y^{(k')}, y^{(k'-1)})\right) A_{k'-1}(y^{(k'-1)}) B_{k'}(y^{(k')})$$

動的計画法による分子の計算（続き）

$$B_\tau(y) = \sum_{y^{(\tau+1)},\ldots,y^{(m_i)}=1}^{c} \exp\left(\sum_{k=\tau+2}^{m_i} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)}) + \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(\tau+1)}, y^{(\tau+1)}, y) \right)$$

- $B_\tau(y)$ は以下のように再帰表現できる 証明は宿題

$$B_\tau(y^{(\tau)}) = \sum_{y^{(\tau+1)}=1}^{c} B_{\tau+1}(y^{(\tau+1)}) \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(\tau+1)}, y^{(\tau+1)}, y^{(\tau)}) \right)$$

$B_\tau(y)$	$\tau = 1$	$\tau = 2$	$\cdots$	$\tau = m$
$y = 1$				
$\vdots$				
$y = c$				

動的計画法による勾配の計算

- $A_1(y^{(1)}), \ldots, A_{m_i}(y^{(m_i)})$ の順に計算
- $B_{m_i}(y^{(m_i)}), \ldots, B_1(y^{(1)})$ の順に計算
- これらを用いて勾配を効率良く計算:

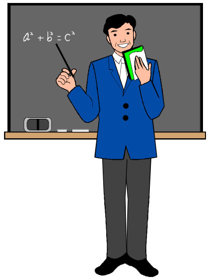
$\mathcal{O}(c^{m_i}) \longrightarrow \mathcal{O}(c^2 m_i)$

"前向き後向きアルゴリズム"

●分母: $\displaystyle\sum_{y^{(m_i)}=1}^{c} A_{m_i}(y^{(m_i)})$

●分子: $\displaystyle\sum_{k'=1}^{m_i} \sum_{y^{(k'-1)}, y^{(k')}=1}^{c} \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k')}, y^{(k')}, y^{(k'-1)})$

$\displaystyle\times \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k')}, y^{(k')}, y^{(k'-1)})\right) A_{k'-1}(y^{(k'-1)}) B_{k'}(y^{(k')})$



講義の流れ

1. 確率的分類（10章）
2. 系列データの分類（11章）
   A) 条件付き確率場モデル
   B) 条件付き確率場モデルの学習
   C) 条件付き確率場モデルによるラベル予測

条件付き確率場による予測

- 事後確率が最大のラベル系列を求める

$$\operatorname*{argmax}_{y^{(1)},\ldots,y^{(m)}\in\{1,\ldots,c\}} q(\overline{y}|\overline{\boldsymbol{x}};\widehat{\zeta})$$

$$q(\overline{y}|\overline{\boldsymbol{x}};\zeta)=\frac{\exp\left(\zeta^{\top}\boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y})\right)}{\sum_{\overline{y}'=1}^{\overline{c}}\exp\left(\zeta^{\top}\boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y}')\right)}$$

- 式の簡略化

$$\operatorname*{argmax}_{y^{(1)},\ldots,y^{(m)}\in\{1,\ldots,c\}} \frac{\exp\left(\widehat{\zeta}^{\top}\boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y})\right)}{\sum_{\overline{y}'=1}^{\overline{c}}\exp\left(\widehat{\zeta}^{\top}\boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y}')\right)}$$

$$=\operatorname*{argmax}_{y^{(1)},\ldots,y^{(m)}\in\{1,\ldots,c\}} \widehat{\zeta}^{\top}\boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y})$$

- $\operatorname*{argmax}_{y^{(1)},\ldots,y^{(m)}\in\{1,\ldots,c\}}$ の計算に $\mathcal{O}(c^m)$ 時間かかる

  $\rightarrow$ 動的計画法で計算

動的計画法による最大化の計算

$$\operatorname*{argmax}_{y^{(1)},\ldots,y^{(m)}\in\{1,\ldots,c\}} \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y})$$

- $y^{(1)},\ldots,y^{(m-1)}$ と $y^{(m)}$ に分解:

$$\max_{y^{(1)},\ldots,y^{(m)}\in\{1,\ldots,c\}} \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y}) = \max_{y^{(m)}\in\{1,\ldots,c\}} P_m(y^{(m)})$$

$$P_\tau(y) = \max_{y^{(1)},\ldots,y^{(\tau-1)}\in\{1,\ldots,c\}} \left[\sum_{k=1}^{\tau-1} \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\boldsymbol{x}^{(k)},y^{(k)},y^{(k-1)}) + \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\boldsymbol{x}^{(\tau)},y,y^{(\tau-1)})\right]$$

- 再帰表現:

証明は宿題

$$P_\tau(y^{(\tau)}) = \max_{y^{(\tau-1)}\in\{1,\ldots,c\}} \left[P_{\tau-1}(y^{(\tau-1)}) + \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\boldsymbol{x}^{(\tau)},y^{(\tau)},y^{(\tau-1)})\right]$$

動的計画法による最大化の計算（続き）

$$\underset{y^{(1)},\ldots,y^{(m)}\in\{1,\ldots,c\}}{\mathrm{argmax}} \quad q(\overline{y}|\overline{\boldsymbol{x}};\widehat{\boldsymbol{\zeta}})$$

$$q(\overline{y}|\overline{\boldsymbol{x}};\boldsymbol{\zeta}) = \frac{\exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y})\right)}{\sum_{\overline{y}'=1}^{\overline{c}} \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\overline{\boldsymbol{x}},\overline{y}')\right)}$$

$$P_\tau(y^{(\tau)}) = \max_{y^{(\tau-1)}\in\{1,\ldots,c\}} \left[P_{\tau-1}(y^{(\tau-1)}) + \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\boldsymbol{x}^{(\tau)}, y^{(\tau)}, y^{(\tau-1)}) \right]$$

- $P_1(y^{(1)}), \ldots, P_m(y^{(m)})$ の順に計算
- これらを用いて最大化を効率良く計算：

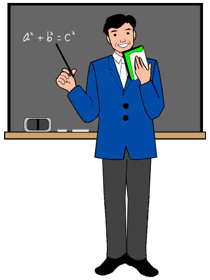
$$\max_{y^{(k)}\in\{1,\ldots,c\}} P_k(y^{(k)}), \quad k = 1, \ldots, m$$

$$\mathcal{O}(c^m) \longrightarrow \mathcal{O}(c^2 m)$$

“Viterbiアルゴリズム”

系列データの分類:まとめ

- 系列データの分類:
  - 個別に分類すると系列としての性質を生かせない
  - ナイーブに計算すると系列の長さの指数時間かかる
- 条件付き確率場:
  - 隣り合ったデータのみの関連を考慮
  - 動的計画法により効率良く計算できる
  - 系列データ以外にも拡張できる



講義の流れ

1. 確率的分類（10章）
2. 系列データの分類（11章）

まとめ

確率的分類

- クラス事後確率 $p(y=\widehat{y}|\boldsymbol{x})$ の学習に基づく分類

$$\widehat{y}=\underset{y=1,\ldots,c}{\operatorname{argmax}}\, p(y|\boldsymbol{x})$$

- クラス予測に対する信頼度が得られる

系列データの分類

- 動的計画法を用いることにより，学習・予測を高速化
- 一般のグラフ構造に拡張できる
- マージン最大化原理に基づいた構造サポートベクトルマシンも提案されている

次回の予告

- オンライン学習（15章）

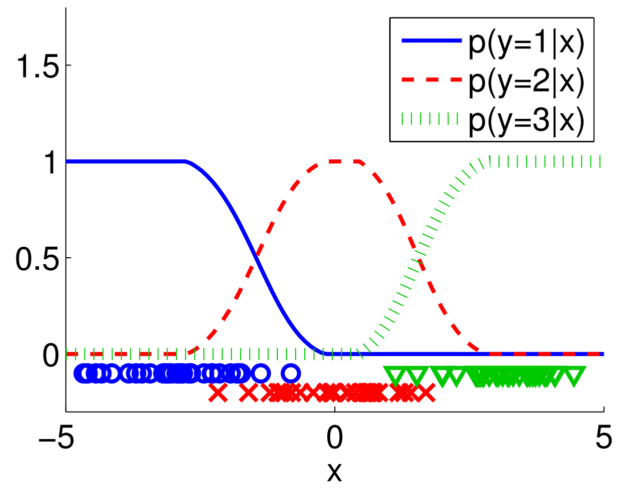
宿題1

■ガウスカーネルモデル

$$q(y|\boldsymbol{x};\boldsymbol{\theta}^{(y)}) = \sum_{j:y_j=y} \theta_j^{(y)} \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_j\|^2}{2h^2}\right)$$

に対して，最小二乗確率的分類を実装せよ

```
n=90; c=3; y=ones(n/c,1)*[1:c]; y=y(:);
x=randn(n/c,c)+repmat(linspace(-3,3,c),n/c,1); x=x(:);
```



宿題2

$$B_\tau(y) = \sum_{y^{(\tau+1)},\ldots,y^{(m_i)}=1}^{c} \exp\left(\sum_{k=\tau+2}^{m_i} \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(k)}, y^{(k)}, y^{(k-1)}) + \boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(\tau+1)}, y^{(\tau+1)}, y) \right)$$

- $B_\tau(y)$ は以下のように再帰表現できることを示せ

$$B_\tau(y^{(\tau)}) = \sum_{y^{(\tau+1)}=1}^{c} B_{\tau+1}(y^{(\tau+1)}) \exp\left(\boldsymbol{\zeta}^\top \boldsymbol{\varphi}(\boldsymbol{x}_i^{(\tau+1)}, y^{(\tau+1)}, y^{(\tau)}) \right)$$

$B_\tau(y)$	$\tau=1$	$\tau=2$	$\cdots$	$\tau=m$
$y=1$				
$\vdots$				
$y=c$				

宿題3

$$P_\tau(y) = \max_{y^{(1)},\ldots,y^{(\tau-1)}\in\{1,\ldots,c\}} \left[\sum_{k=1}^{\tau-1} \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\boldsymbol{x}^{(k)}, y^{(k)}, y^{(k-1)}) + \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\boldsymbol{x}^{(\tau)}, y, y^{(\tau-1)}) \right]$$

- 以下の再帰表現を求めよ

$$P_\tau(y^{(\tau)}) = \max_{y^{(\tau-1)}\in\{1,\ldots,c\}} \left[P_{\tau-1}(y^{(\tau-1)}) + \widehat{\boldsymbol{\zeta}}^\top \boldsymbol{\varphi}(\boldsymbol{x}^{(\tau)}, y^{(\tau)}, y^{(\tau-1)}) \right]$$

$P_\tau(y)$	$\tau = 1$	$\tau = 2$	$\cdots$	$\tau = m$
$y = 1$				
$\vdots$				
$y = c$				